Unsupervised Functional Data Clustering Based on Adaptive Weights

Yutong Gao[1], Shuang Chen[2*]

[1]School of Sciences, Hebei University of Technology, Tianjin, China
[2]School of Sciences, Beijing Institute of Petrochemical Technology, Beijing, China
Email: *shchen@bjpt.edu.cn

Abstract

In recent years, functional data has been widely used in finance, medicine, biology and other fields. The current clustering analysis can solve the problems in finite-dimensional space, but it is difficult to be directly used for the clustering of functional data. In this paper, we propose a new unsupervised clustering algorithm based on adaptive weights. In the absence of initialization parameter, we use entropy-type penalty terms and fuzzy partition matrix to find the optimal number of clusters. At the same time, we introduce a measure based on adaptive weights to reflect the difference in information content between different clustering metrics. Simulation experiments show that the proposed algorithm has higher purity than some algorithms.

Keywords

Functional Data, Unsupervised Learning Clustering, Functional Principal Component Analysis, Adaptive Weight

1. Introduction

Functional data analysis (FDA), proposed by J.O. Ramsay [1], is a high-dimensional data analysis method based on the functionalization of discrete data to describe statistical relationships from a continuous perspective. It is characterized by treating discrete data as a functional whole with an inherently uniform structure, rather than just an arrangement of individual observations. In recent years, the extension of classical multivariate statistical methods to functional data is an important part of data mining research. In the scalar environment, there are many mature multivariate clustering algorithms, such as Affinity Propagation clustering, hierarchical clustering, K-means clustering, DBSCAN clustering, Chameleon clustering, etc. Most of the multivariate cluster analysis methods

cannot be directly applied to functional data. The intuitive analysis idea is to discretize functional data based on sampling, and then use multivariate clustering model to cluster discrete data points [2] [3]. Although this can solve the problem that functional data is difficult to apply multivariate clustering algorithms, it also violates the main idea of "taking the whole function as the analysis object". Tarpey T *et al.* [4] proposed a clustering algorithm based on function principal points. The clustering results of this algorithm are affected by the number and location of extreme points, therefore it lacks robustness. Tokushige [5] extended the K-means clustering algorithm to functional data, where the cluster center is determined by the minimum value of the objective function. Antoniadis *et al.* [6] used wavelet analysis techniques in functional data clustering. Wang *et al.* [7] proposed a weighted clustering algorithm based on functional principal components, which can effectively make up for the limitations of ordinary functional principal component clustering analysis. Delaigle *et al.* [8] proposed a weighted K-means algorithm by using the Haar base function proposed by Hardle [9] and the unbalanced Haar base function proposed by Fryzlewicz [10]. This algorithm selects the projection function by weighting the K-means clustering criteria and proves that nearly perfect clustering can be achieved by projecting the data onto the chosen finite-dimensional space. Sharp and Browne [11] imposed jointly generalized hyperbolic distributions on projections of basis expansion coefficients into group specific subspaces. Parameter estimation is done through a multicycle ECM algorithm. Wu *et al.* [12] assumed that the distribution of the cluster members around their projections onto the cluster's principal curve is an isotropic Gaussian, and that the projections of cluster members onto the cluster's principal curve are uniformly distributed. The proposed method employe Bayesian Information Criterion to automatically and simultaneously find the appropriate number of features, the optimal degree of smoothing and the corresponding cluster members. It can be noticed that cluster analysis based on continuous perspective and cluster analysis of discrete data still have a lot in common. If infinite-dimensional functional clustering can be transformed into finite-dimensional clustering, then many excellent multivariate clustering analysis can be used in functional data clustering.

2. Preliminaries

2.1. K-Means

The K-means is one of the most popular unsupervised learning algorithms that solve the well-known clustering problem. Let $X = \{x_1, \cdots, x_n\}$ be a data set in a d-dimensional Euclidean space R^d. Let $V = \{v_1, \cdots, v_c\}$ be the cluster centers. Let $Z = [z_{ik}]_{n \times c}$, where $z_{ik} \in \{0,1\}$ is binary variable, indicating if the data point x_i belongs to k-th cluster, $k = 1, 2, \cdots, c$. The K-means objective function is

$$J(z, V) = \sum_{i=1}^{n} \sum_{k=1}^{c} z_{ik} \|x_i - v_k\|^2 \tag{1}$$

The K-means algorithm is iterated through necessary conditions for minimizing the K-means objective function $J(z, V)$ with updating equations for cluster centers and memberships, respectively, as

$$v_k = \frac{\sum_{i=1}^{n} z_{ik} x_i}{\sum_{i=1}^{n} z_{ik}} \tag{2}$$

$$z_{ik} = \begin{cases} 1 & \text{if } \|x_i - v_k\|^2 = \min_{1 \leq k \leq c} \|x_i - v_k\|^2 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where $\|x_i - v_k\|$ is the Euclidean distance between the data point x_i and the cluster center v_k.

2.2. Functional Principal Component Analysis

Functional principal component analysis (FPCA) is a generalization of principal component analysis. Suppose $x_i(t)$ is a random process defined on a compact interval *T*. Let $\mu(t)$ be the mean function. Let $R(s,t) = \text{cov}\{x(s), x(t)\}$ be the covariance function. By using the spectral decomposition of the covariance function, we can obtain:

$$R(s,t) = \sum_{p=1}^{\infty} \lambda_p \varphi_p(s) \varphi_p(t)$$

where λ_p is a non-negative eigenvalue of $R(s,t)$ and satisfies $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$. $\varphi_p(s)$ is the corresponding eigenfunction. The eigenfunctions are regarded as a set of basis functions, called principal component basis functions. The principal component basis function combined with the Karhunen-Loeve expansion can represent the infinite-dimensional function data in a limited manner. Project X into the principal component basis function, and the principal component score of the function ξ_p is the corresponding coefficient.

$$\xi_p = \int_\tau X(t) \varphi_p(t) \, dt \tag{4}$$

Applying the Karhunen-Loeve expansion, we can get

$$X(t) = \mu(t) + \sum_{p=1}^{\infty} \xi_p \varphi_p(t) \tag{5}$$

3. Unsupervised Functional Data Clustering Based on Adaptive Weights

Multivariate clustering analysis is difficult to directly extend to functional data clustering. But from the knowledge of principal component analysis in the previous section, functional data $X(t)$ is equivalent to its principal component score $\xi_i = (\xi_{i1}, \cdots, \xi_{id})^T$. Since the variance contribution rates of the principal component scores are different, the importance of different principal components is obviously different. It is difficult to achieve an ideal classification effect,

if we ignore the objective differences in the efficiency of different principal component scores, and use the principal component score to represent the sample function for clustering [13]. K-means cluster analysis is not completely unsupervised learning, because it requires the number of clusters to be given first. Therefore, we use functional principal component analysis to map functional data from infinite-dimensional space to finite-dimensional space, and then use unsupervised functional data clustering based on adaptive weights (UFDCAW) to cluster $\xi_i = \left(\xi_{i1}, \cdots, \xi_{id}\right)^{\mathrm{T}}$ on the finite-dimensional space. Finally, the clustering result of the functional data $X(t)$ is obtained.

Let $\xi = \{\xi_1, \cdots, \xi_n\}$ which is equivalent to $X(t) = \{X_1(t), \cdots, X_n(t)\}$ be the data on the d-dimensional Euclidean space R^d. Let $V = \{v_1, \cdots, v_c\}$ be the set of c class centers. Let $u = \{u_{ik}\}_{n \times c}$ be the degree of membership function matrix, where u_{ik} is the membership degree of ξ_i belonging to the k class. For any i, k, there is $0 \le u_{ik} \le 1$ and $\sum_{k=1}^{c} u_{ik} = 1$. Let $a = (\alpha_1, \cdots, \alpha_c)$ be a set of mixed proportions, where α_k represents the probability that data point belongs to the k class, and satisfies $\sum_{k=1}^{c} \alpha_k = 1$. $-\sum_{k=1}^{c} \alpha_k \ln \alpha_k$ is the entropy. The objective function of unsupervised functional data clustering based on adaptive weights is

$$J(u, V, \alpha) = \sum_{i=1}^{n} \sum_{k=1}^{c} u_{ik}^m d_{ik}^2 - n\beta \sum_{k=1}^{c} \alpha_k \ln \alpha_k - \gamma \sum_{i=1}^{n} \sum_{k=1}^{c} u_{ik} \ln \alpha_k \qquad (6)$$

where d_{ik} is the distance from the data point ξ_i to the class center v_k. Considering the difference in the scoring efficiency of different principal components, let

$$d_{ik} = \sqrt{\sum_{p=1}^{d} \left(\omega_p \left(\xi_{ip} - v_{kp} \right)^2 \right)}$$

where $\omega_p = \lambda_p \Big/ \sum_{p=1}^{d} \lambda_p$ is the adaptive weight of the p item of the principal component score.

The Lagrangian function of formula (3.1) is

$$\tilde{J}(u, V, \alpha, r_1, r_2) = \sum_{i=1}^{n} \sum_{k=1}^{c} u_{ik}^m d_{ik}^2 - n\beta \sum_{k=1}^{c} \alpha_k \ln \alpha_k - \gamma \sum_{i=1}^{n} \sum_{k=1}^{c} u_{ik} \ln \alpha_k$$
$$- r_1 \left(\sum_{k=1}^{c} \alpha_k - 1 \right) - r_2 \left(\sum_{k=1}^{c} u_{ik} - 1 \right) \qquad (7)$$

First, we take the partial derivative of Largrangian (3.2) with respect to u_{ik}, we obtain

$$\frac{\partial \tilde{J}(u, V, \alpha, r_1, r_2)}{\partial u_{ik}} = m u_{ik}^{m-1} d_{ik}^2 - \gamma \ln \alpha_k - r_2 = 0 \qquad (8)$$

Thus, we have

$$u_{ik} = \left(\frac{\gamma \ln \alpha_k + r_2}{m} \right)^{\frac{1}{m-1}} d_{ik}^{-\frac{2}{m-1}}$$

From $\sum_{k=1}^{c} u_{ik} = 1$, we can get

$$u_{ik} = \frac{1}{\sum_{s=1}^{c} \left(d_{ik}/d_{is} \right)^{\frac{2}{m-1}}} \tag{9}$$

Then we get the update equation for the center point v_k as follows:

$$v_k = \frac{\sum_{i=1}^{n} u_{ik}^{m} \xi_i}{\sum_{i=1}^{n} u_{ik}^{m}} \tag{10}$$

We next take the partial derivative of the Lagrangian (3.2) with respect to α_k

$$\frac{\partial \tilde{J}(u, V, \alpha, r_1, r_2)}{\partial \alpha_k} = -n\beta \left(\ln \alpha_k + 1 \right) - \gamma \sum_{i=1}^{n} \frac{u_{ik}}{\alpha_k} - r_1 = 0 \tag{11}$$

then we get the updating equation for α_k as follows:

$$\alpha_k^{(t+1)} = \sum_{i=1}^{n} \frac{u_{ik}}{n} + \frac{\beta}{\gamma} \alpha_k^{(t)} \left(\ln \alpha_k^{(t)} - \sum_{s=1}^{c} \alpha_s^{(t)} \ln \alpha_s^{(t)} \right) \tag{12}$$

where t is the number of iterations of the algorithm.

$\sum_{s=1}^{c} \alpha_s \ln \alpha_s$ is the weighted mean of $\ln \alpha_k$ in equation (3.7). For the k-th mixing proportion $\alpha_k^{(t)}$, if $\ln \alpha_k^{(t)}$ is less than the weighted mean, then the new $\alpha_k^{(t+1)}$ will be less than the $\alpha_k^{(t)}$. That is, the smaller proportion will decrease, and then competition will occur. If $\alpha_k \leq 0$ or $\alpha_k < \frac{1}{n}$ for some k, they are considered to be illegitimate proportions. In this situation, we discard those clusters and then update the cluster number $c^{(t)}$ to be

$$c^{(t+1)} = c^{(t)} - \left| \left\{ \alpha_k^{(t+1)} \middle| \alpha_k^{(t+1)} < \frac{1}{n}, k = 1, \cdots, c^{(t)} \right\} \right| \tag{13}$$

where $\left| \{ \} \right|$ denotes the cardinality of the set $\{ \}$. After updating the number of clusters $c^{(t+1)}$, the remaining mixing proportion $\alpha_k^{(t+1)}$ and corresponding $u_{ik}^{(t+1)}$ need to be re-normalized by

$$\alpha_k^* = \frac{\alpha_k^*}{\sum_{s=1}^{c^{(t+1)}} \alpha_s^*} \tag{14}$$

$$u_{ik}^* = \frac{u_{ik}^*}{\sum_{s=1}^{c^{(t+1)}} u_{is}^*} \tag{15}$$

We next concern about the parameter learning of γ and β for the two terms of $\sum_{i=1}^{n} \sum_{k=1}^{c} z_{ik} \ln \alpha_k$ and $\sum_{i=1}^{n} \sum_{k=1}^{c} z_{ik} \ln \alpha_k$. Based on some increasingly learning rates of cluster number with $e^{-c^{(t)}/100}, e^{-c^{(t)}/250}, e^{-c^{(t)}/500}, e^{-c^{(t)}/750}$, and $e^{-c^{(t)}/1000}$, it is seen that $e^{-c^{(t)}/100}$ decreases faster, but $e^{-c^{(t)}/500}, e^{-c^{(t)}/750}, e^{-c^{(t)}/1000}$ decreases slower [14]. We suppose that the parameter γ should not decrease too slow or

too fast, and so we set the parameter γ as

$$\gamma^{(t)} = e^{-c^{(t)}/250} \tag{16}$$

To guarantee condition $\max_{1 \le k \le c} \alpha_k^{(t+1)} \le 1$, the parameters β are defined as

$$\beta^{(t+1)} = \min \left(\frac{\sum_{k=1}^{c} \exp\left(-n\eta \left| \alpha_k^{(t+1)} - \alpha_k^{(t)} \right| \right)}{c}, \frac{\gamma \left(1 - \sum_{i=1}^{n} \frac{u_{ik}}{n}\right)}{-\max_{1 \le k \le c} \alpha_k^{(t)} \left(\sum_{s=1}^{c} \alpha_s^{(t)} \ln \alpha_s^{(t)} \right)} \right) \tag{17}$$

where $\eta = 2 / d^{\lfloor d/2-1 \rfloor}$ and $\lfloor a \rfloor$ represents the largest integer that is no more than a and t denotes the iteration number in the algorithm. In Equation (3.12), if $\left| \alpha_k^{(t+1)} - \alpha_k^{t} \right|$ is large, then β is smalll to maintain stability. Although functional principal component analysis expands the function data from infinite dimension to finite dimension, the dimension of the sample is still relatively high, so we add $\eta = 2 / d^{\lfloor d/2-1 \rfloor}$ adjust parameter β to Equation (3.12).

In our setting, we use all data points as initial means with $v_k = \xi_k$, *i.e.* $c^{(0)} = n$, and we use $\alpha_k^{(0)} = 1/n$ as initial mixing proportions. Thus, the proposed unsupervised Learning functional data clustering based on adaptive weights can be summarized as **Algorithm 1**.

Algorithm 1. Unsupervised Learning Functional Data Clustering Based on Adaptive Weights

1: Calculate the principal component score of function data $\xi_i, i = 1, \ldots, n$.
2: Fix $\varepsilon > 0$. Give initial $c^{(0)} = n, \alpha_k^{(0)} = \frac{1}{n}, v_k^{(0)} = \xi_i$, and initial learning rates $\gamma^{(0)} = \beta^{(0)} = 1$. Set $t = 0$.
3: Compute $\mu_{ik}^{(t+1)}$ using $v_k^{(t)}$ by Eq.(9).
4: Compute $\gamma^{(t+1)}$ by Eq.(16).
5: Update $\alpha_k^{(t+1)}$ with $\mu_{ik}^{(t+1)}$ and $\alpha_k^{(t)}$ by Eq.(12).
6: Compute $\beta^{(t+1)}$ with $\alpha_k^{(t)}$ and $\alpha_k^{(t+1)}$ by Eq.(17).
7: Update $c^{(t)}$ to $c^{(t+1)}$ by discard those clusters and adjust $\alpha_k^{(t+1)}$ and $\mu_{ik}^{(t+1)}$ by Eq.(14) and Eq.(15).
8: Update $v_k^{(t+1)}$ by Eq.(10).
9: Compare $v_k^{(t+1)}$ and $v_k^{(t)}$. If $\sqrt{\sum_{p=1}^{d} \left(\omega_p (v_{kp}^{(t+1)} - v_{kp}^{(t)})^2 \right)} < \varepsilon$, then stop. Else $t = t + 1$ and return to step 3.

---

4. Theoretical Properties

Proposition 1. Given the iterative formula $\alpha_k^{(t)}$ that minimizes the objective function $J(u, V, \alpha) = \sum_{i=1}^{n} \sum_{k=1}^{c} u_{ik}^m d_{ik}^2 - n\beta \sum_{k=1}^{c} \alpha_k \ln \alpha_k - \gamma \sum_{i=1}^{n} \sum_{k=1}^{c} u_{ik} \ln \alpha_k$, if there exists k such that $\ln \alpha_k^{(t)}$ is less than its weighted mean, then $\alpha_k^{(t+1)} < \alpha_k^{(t)}$.

proof.

$$\alpha_k^{(t+1)} = \sum_{i=1}^{n} \frac{u_{ik}}{n} + \frac{\beta}{\gamma} \alpha_k^{(t)} \left(\ln \alpha_k^{(t)} - \sum_{s=1}^{c} \alpha_s^{(t)} \ln \alpha_s^{(t)} \right) < \sum_{i=1}^{n} \frac{u_{ik}}{n} + \frac{\beta}{\gamma} \alpha_k^{(t)} \left(\ln \alpha_k^{(t)} - \ln \alpha_k^{(t)} \right)$$

$$= \alpha_k^{(t)}$$

That is, there exists k such that $\ln \alpha_k^{(t)} < \sum_{s=1}^{c} \alpha_s^{(t)} \ln \alpha_s^{(t)}$ is satisfied, we can get

$\alpha_k^{(t+1)} < \alpha_k^{(t)}$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proposition 2. Given the iterative formula $\alpha_k^{(t)}$ that minimizes the objective function $J(u,V,\alpha) = \sum_{i=1}^{n}\sum_{k=1}^{c} u_{ik}^m d_{ik}^2 - n\beta \sum_{k=1}^{c} \alpha_k \ln \alpha_k - \gamma \sum_{i=1}^{n}\sum_{k=1}^{c} u_{ik} \ln \alpha_k$, if

$\beta \le -\gamma \left(1 - \sum_{i=1}^{n} \dfrac{u_{ik}}{n}\right) \Big/ \max_{1\le k\le c} \alpha_k^{(t)} \left(\sum_{s=1}^{c} \alpha_s^{(t)} \ln \alpha_s^{(t)}\right)$, then $\max_{1\le k\le c} \alpha_k^{(t+1)} \le 1$.

proof.

$$
\begin{aligned}
\max_{1\le k\le c} \alpha_k^{(t+1)} &= \max_{1\le k\le c} \sum_{i=1}^{n} \frac{u_{ik}}{n} + \frac{\beta}{\gamma} \max_{1\le k\le c} \alpha_k^{(t)} \left(\ln \max_{1\le k\le c} \alpha_k^{(t)} - \sum_{s=1}^{c} \alpha_s^{(t)} \ln \alpha_s^{(t)} \right) \\
&< \max_{1\le k\le c} \sum_{i=1}^{n} \frac{u_{ik}}{n} + \frac{\beta}{\gamma} \left(-\max_{1\le k\le c} \alpha_k^{(t)} \left(\sum_{s=1}^{c} \alpha_s^{(t)} \ln \alpha_s^{(t)} \right) \right) \\
&< \frac{1}{n} \max_{1\le k\le c} \sum_{i=1}^{n} u_{ik} - \frac{ \gamma \left(1 - \sum_{i=1}^{n} \dfrac{u_{ik}}{n}\right) \left(-\max_{1\le k\le c} \alpha_k^{(t)} \left(\sum_{s=1}^{c} \alpha_s^{(t)} \ln \alpha_s^{(t)} \right) \right) }{ \max_{1\le k\le c} \alpha_k^{(t)} \left(\sum_{s=1}^{c} \alpha_s^{(t)} \ln \alpha_s^{(t)} \right) \gamma } \\
&= 1
\end{aligned}
$$

Therefore, when

$$
\beta^{(t+1)} = \min \left(\frac{\sum_{k=1}^{c} \exp\left(-n\eta \left| \alpha_k^{(t+1)} - \alpha_k^{(t)} \right|\right)}{c}, \frac{\gamma \left(1 - \sum_{i=1}^{n} \dfrac{u_{ik}}{n}\right)}{-\max_{1\le k\le c} \alpha_k^{(t)} \left(\sum_{s=1}^{c} \alpha_s^{(t)} \ln \alpha_s^{(t)} \right)} \right)
$$

holds, $\max_{1\le k\le c} \alpha_k^{(t+1)} \le 1$ can be guaranteed. $\qquad\qquad\qquad\square$

5. Simulation Study

In this section, we use simulation experiments to verify the performance of the proposed clustering algorithm and whether it can calculate the optimal number of clusters c. At the same time, we compare with fuzzy C-means (FCM) and adaptive weighting functional clustering (AWFC). The result is evaluated using cluster purity. Purity is calculated using Equation (5.1):

$$
\text{purity}(\Omega, C) = \frac{1}{N} \sum_{K=1}^{K} \max_{1\le j\le J} \left| w_k \cap c_j \right| \tag{18}
$$

where $\Omega = \{w_1, \cdots, w_K\}$ is the cluster division, $C = \{c_1, \cdots, c_J\}$ is the real class division. The value range of purity is $[0,1]$, the larger the value, the better the clustering performance.

Assume that the three functional data for the simulation experiment are:
$X_1(t) = t + t^3 / \sqrt{5}$, $X_2(t) = t + \sin(\pi t)/\sqrt{5}$, $X_3(t) = t + (t-2)^3 / \sqrt{5}$, $t \in [0,1]$. $Y_{ij} = X_i(t_{ij}) + \varepsilon_{ij}$ is the observation value of the function data with error, and the error ε_{ij} obeys a normal distribution with mean 0 and variance 0.02. In order to ensure that the FCM clustering algorithm is available when simulating expe-

riments, we assume that t_{ij} is 100 equally spaced observation nodes on the interval $[0,1]$. The observation sample size of each function data in the simulation experiment is the same, and the total sample size is 30, 300, 600, in order to compare the effectiveness of the clustering algorithm under different sample sizes. The figure shows the clustering data when $n = 300$. **Figure 1(a)** is the observation value with noise. **Figure 1(b)** is the function data generated using the B-spline basis.

We can see that the three types of curves are increasing functions from **Figure 1(a)** and **Figure 1(b)**, but there are significant differences in their derivatives. Following the steps of UFDCAW, the scatter plot of functional principal component scores of 3rd and 5th iterations for a sample size of 300 is shown in **Figure 2**. From **Figure 2(b)**, we can see that the number of clusters decreases rapidly from 300 to 165 after 3 iterations. The algorithm ends after five iterations and obtains the correct number of clusters $c = 3$ and three class centers. The purity of clustering is 0.993.

Table 1 shows the average of the results of 100 replicate experiments using FCM, AWFC and UFDCAW on simulated data with sample sizes of 30, 300 and 600, judged by the purity of the clustering results.

The data presented in **Table 1** shows that UFDCAW has higher cluster purity than the FCM and AWFC for sample sizes of 30, 300 and 600. The results of the comparison show that: 1) UFDCAW has higher cluster purity than FCM and can display sample distribution characteristics in low dimensional space. 2) FCM and AWFC required random initial class centres, therefore the cluster prurity will be low when algorithm selects an unsuitable class center. UFDCAW is more

Table 1. Comparison of clustering algorithm results.

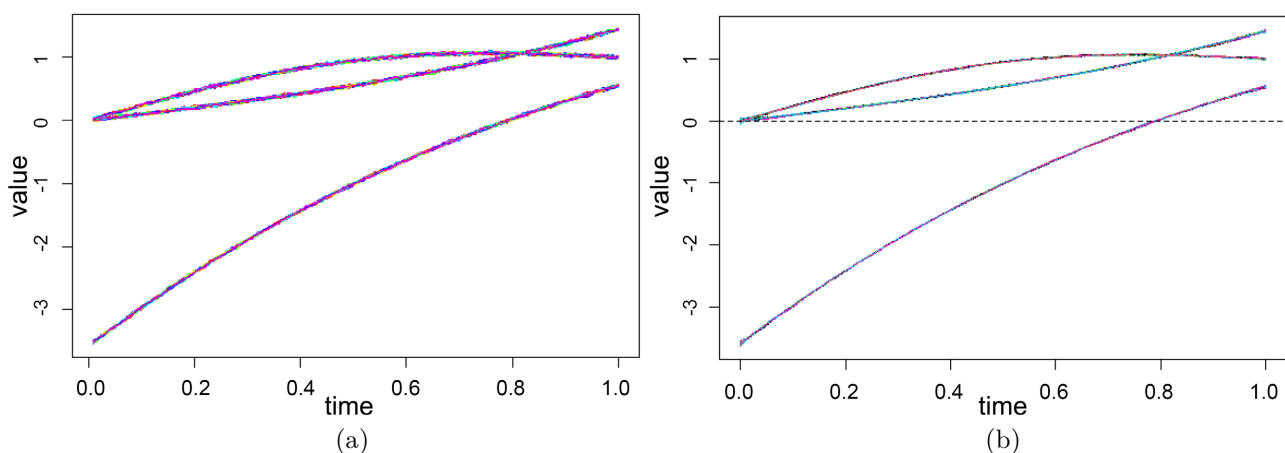|	n	FCM	AWFC	UFDCAW
	30	0.839	0.801	1
Purity	300	0.825	0.856	0.993
	600	0.805	0.871	0.997

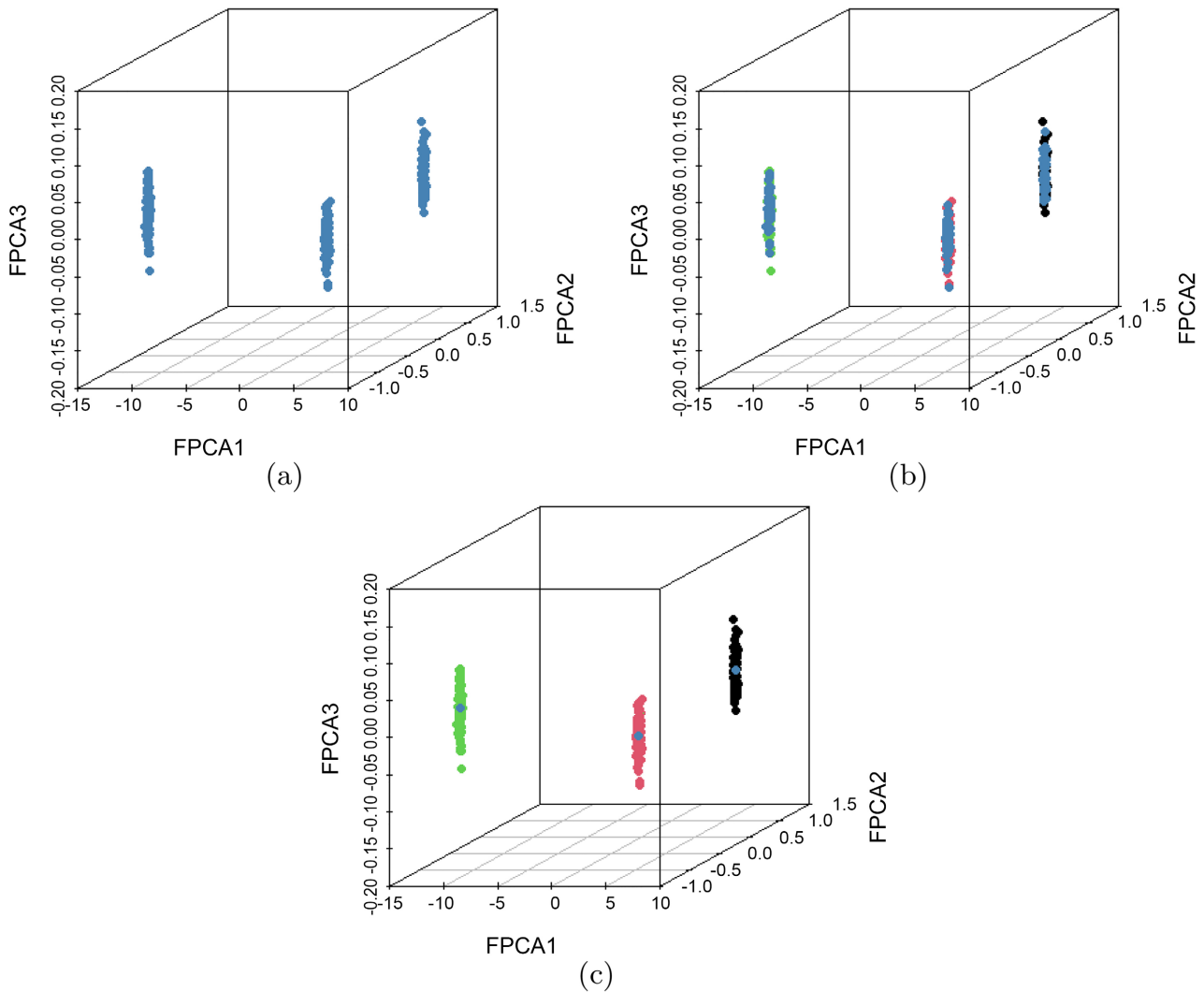**Figure 1.** (a) Observation value with noise, (b) function data for a sample size of 30.

Figure 2. Scatter plot of the functional principal component scores of data with a sample size of 30 (a) $t = 0$, (b) $t = 3$, (c) $t = 5$.

stable because there is no need to select the initial clustering center. 3) UFDCAW can be iterated to obtain the number of clusters.

6. Conclusion

At present, the data we collect is becoming more and more complicated. Using multivariate statistical models to deal with complicated data will cause problems such as dimensional disaster, and functional data emerges as the times require. Based on the former studies, we expand the K-means clustering algorithm based on adaptive weights. It not only reduces the influence of noise, but also can find the optimal number of clusters using the algorithm when the initialization parameters are missing. It is verified by simulation that the algorithm proposed in this paper can obtain more accurate clustering results and has a certain degree of feasibility. But, the algorithm is slightly more complicated, and some information will be lost when using basis function fitting and functional principal component analysis, which needs to be further improved.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Ramsay, J.O. (1982) When the Data Are Functions. *Psychometrika*, **47**, 379-396. https://doi.org/10.1007/BF02293704

[2] Bouveyron, C., Girard, S. and Schmid, C. (2007) High-Dimensional Data Clustering. *Computational Statistics & Data Analysis*, **52**, 502-519. https://doi.org/10.1016/j.csda.2007.02.009

[3] Zhu, J. and Chen, K. (2007) The Cluster Analysis of Panel Data and Its Application. *Statistical Research*, **24**, 11-14.

[4] Tarpey, T. and Kinateder, K.K.J. (2003) Clustering Functional Data. *Journal of Classification*, **20**, 93-114. https://doi.org/10.1007/s00357-003-0007-3

[5] Tokushige, S., Yadohisa, H. and Inada, K. (2007) Crisp and Fuzzy K-Means Clustering Algorithms for Multivariate Functional Data. *Computational Statistics*, **22**, 1-16. https://doi.org/10.1007/s00180-006-0013-0

[6] Antoniadis, A., Brossat, X., Cugliari, J. and Poggi, J.M. (2013) Clustering Functional Data Using Wavelets. *International Journal of Wavelets, Multiresolution and Information Processing*, **11**, 1350003. https://doi.org/10.1142/S0219691313500033

[7] Wang, D., Zhu, J. and Wang, D. (2015) Research of Clustering Analysis for Functional Data Based on Adaptive Weighting. *Journal of Applied Statistics and Management*, **34**, 84-92.

[8] Delaigle, A., Hall, P. and Pham, T. (2019) Clustering Functional Data into Groups by Using Projections. *Journal of the Royal Statistical Society: Series B: Statistical Methodology*, **81**, 271-304. https://doi.org/10.1111/rssb.12310

[9] Hardle, W., Kerkyacharian, G., Picard, D. and Tsybakov, A. (2012) Wavelets, Approximation, and Statistical Applications. Springer Science & Business Media, Berlin.

[10] Fryzlewicz, P. (2007) Unbalanced Haar Technique for Nonparametric Function Estimation. *Journal of the American Statistical Association*, **102**, 1318-1327. https://doi.org/10.1198/016214507000000860

[11] Sharp, A. and Browne, R. (2021) Functional Data Clustering by Projection into Latent Generalized Hyperbolic Subspaces. *Advances in Data Analysis and Classification*, **15**, 735-757. https://doi.org/10.1007/s11634-020-00432-5

[12] Wu, R., Wang, B. and Xu, A. (2022) Functional Data Clustering Using Principal Curve Methods. *Communications in Statistics-Theory and Methods*, **51**, 7264-7283. https://doi.org/10.1080/03610926.2021.1872636

[13] Wang, D., Zhu, J. and Xie, B. (2012) Research on Clustering Analysis for Functional Data Based on Adaptive Iteration. *Statistical Research*, **4**, 91-96.

[14] Sinaga, K.P. and Yang, M.S. (2020) Unsupervised K-Means Clustering Algorithm. *IEEE Access*, **8**, 80716-80727. https://doi.org/10.1109/ACCESS.2020.2988796